

LAMARIE S. SMITH, Texas
CHAIRMAN

R. JAMES HENSENBRENNER, JR., Wisconsin
HOWARD COBLE, North Carolina
ELTON GALLEGLY, California
BOB GOODLATTE, Virginia
DANIEL E. LUDERKE, California
STEVE CHABOT, Ohio
DARRELL E. ISSA, California
MIKE PENDE, Indiana
J. RANDY FORBES, Virginia
STEVE KING, Iowa
TRENT FRANKS, Arizona
LOUIE GOMPERT, Texas
JIM JORDAN, Ohio
TED POE, Texas
JASON CHAFFETZ, Utah
TIM CROFFER, Arkansas
TOM MARINO, Pennsylvania
PREY GOWDY, South Carolina
DENNIS ROUSE, Florida
SANDY ADAMS, Florida
SEB QUAYLE, Arizona
MARK AMODEI, Nevada

JOHN CONYERS, JR., Michigan
RANKING MEMBER

HOWARD L. Berman, California
JERROLD BADLER, New York
ROBERT C. "BOBBY" SCOTT, Virginia
MELVIN L. WATT, North Carolina
JOE LORESEN, California
SHEILA JACKSON LEE, Texas
MAXINE WATERS, California
STEVE COHEN, Tennessee
HERRY C. "HANK" JOHNSON, JR., Georgia
PEDRO R. PIERLUISI, Puerto Rico
MIKE QUIGLEY, Illinois
JUDY CHU, California
TED DEUTCH, Florida
LINDA T. SANCHEZ, California
JARED POLIS, Colorado

ONE HUNDRED TWELFTH CONGRESS

Congress of the United States

House of Representatives

COMMITTEE ON THE JUDICIARY

2138 RAYBURN HOUSE OFFICE BUILDING

WASHINGTON, DC 20515-6216

(202) 225-3951

<http://www.house.gov/judiciary>

May 31, 2012

The Honorable Eric H. Holder, Jr.
Attorney General
U.S. Department of Justice
Washington, D.C. 20530

Dear Attorney General Holder,

I am concerned that the Justice Department's opposition to the Texas voter ID law may rest on flawed statistical analysis.

Voter ID laws protect the rights of lawful voters and help inspire public confidence in our electoral system. That is why the public overwhelmingly supports such laws and the Supreme Court has upheld them.¹ Still, the Justice Department objects and insists Hispanic voters in Texas are less likely to possess photo IDs. However, Census Bureau data suggests that after accounting for the margin of error in DOJ's calculations the claimed disparity disappears.

On May 27, 2011, Governor Perry signed S.B. 14, requiring Texas voters to present either a citizenship certificate or valid photo ID at polling stations. Under the law, voters without a government issued photo ID may obtain one for free. On March 12, 2012, the Justice Department announced it would use its authority under the Voting Rights Act to block this common sense law from taking effect.

Assistant Attorney General Tom Perez told Texas officials "[e]ven using the data most favorable to the state, Hispanics disproportionately lack either a driver's license or a personal identification card . . . and that disparity is statistically significant."² However, a Census Bureau Report suggests that the disparity is not statistically significant because of the difficulties of determining who is Hispanic.

The Justice Department cites numbers showing 93.7% of Hispanic voters in Texas have photo ID as do 95.7% of non-Hispanics. So the disparity Mr. Perez calls "significant" is a difference of just 2%, and this is before the margin of error is taken into account.

¹ Press Release, Rasmussen Reports, 73% Think Photo ID Requirement Before Voting Does Not Discriminate (Apr. 16, 2012), available at

http://www.rasmussenreports.com/public_content/politics/general_politics/april_2012/73_think_photo_id_requirement_before_voting_does_not_discriminate; Crawford v. Marion County Election Board, 553 U.S. 181 (2008).

² Letter from Thomas E. Perez, Assistant Attorney General, Civil Rights Division, to Keith Ingram, Director of Elections, Office of the Texas Secretary of State (Mar. 12, 2012).

2:13-cv-193
09/02/2014

DEF2330

exhibiticker.com

US_00020674

USA_00057246

The Honorable Eric H. Holder, Jr.
May 31, 2012
Page 2

Since Texas did not track its citizens by race until recently, analysts had to guess who was Hispanic based on surname. They matched names in the Texas drivers' license and voter registration databases against a list of common Hispanic surnames from the Census Bureau. This guesswork introduces errors into the calculation that the Census Bureau has quantified.

A national study found that 30.4% of the people flagged by a standard surname list are not in fact Hispanic. Conversely, 26.3% of people who are Hispanic are not identified as such by this method. The state level results show surname lists are more reliable in areas with large Hispanic populations, but the error rate is still high. For Texas, the error rates for the commonly used Passel-Word Spanish Surname list were 4.2% and 11.9% respectively.³


It is difficult to see how the Justice Department can assert that a 2% disparity is statistically significant when the method used to identify individuals as Hispanic is both over- and under-inclusive by margins that are much greater than 2%. Even the 5.9% disparity DOJ found using a second data set less favorable to Texas loses its significance in light of the fact that the Passel surname list undercounts Hispanic individuals by nearly twice that margin.

The unreliability of the surname list method also taints DOJ's other analyses such as the claim that Hispanics are less likely to live close to government offices that issue the free photo IDs. The disparity may or may not be real because it seems to be within the margin of error.

Accordingly, we ask you provide the following information on the Department's methodology by 10 a.m. on June 6, 2012. This is short notice, but this is a simple request for information DOJ should already have.

1. Did DOJ analysts review the attached Census Bureau report or any similar study analyzing the error rate in Hispanic surname analysis before they made their conclusions of a disparity? If so, why was the error rate issue not addressed, at least in a footnote, in the Department's technical letter to Texas?
2. Please provide the Department's technical analysis demonstrating why the high error rate in Hispanic surname analysis would not prejudice the disparity calculation.

Sincerely,



Lamar Smith
Chairman

cc: The Hon. John Conyers, Jr.

Enclosure

³ R. Colby Perkins, *Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census Post Enumeration Survey Results*, (August 1993), available at <http://www.census.gov/population/www/documentation/twps0004.html>.

US_00020675

USA_00057247

U.S. Census Bureau

**Evaluating the Passel-Word Spanish
Surname List: 1990 Decennial Census
Post Enumeration Survey Results**

R. Colby Perkins

Population Division
U.S. Bureau of the Census
Washington, DC 20233

August 1993

Population Division Working Paper No. 4

This paper reports the general results of research undertaken by the Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

This paper was initially presented during the "Immigrants and Migrants Within the United States" contributed paper session of the Summer 1993 American Statistical Association Meetings.

The author would like to thank the following people for their technical expertise and comments concerning this paper: Jennifer Day, Jorge del Pinal, Edward Fernandez, Randy Klear, John Long, Nampeo McKenney, Kenneth Sausman, Signe Wetrogan, and David Word.

If you have any questions concerning this report, please e-mail a message to pop@census.gov. Include the name of this report and author in the body of the message.

**EVALUATING THE PASSEL-WORD SPANISH SURNAME LIST:
1990 DECENNIAL CENSUS POST ENUMERATION SURVEY RESULTS**

INTRODUCTION

The 1990 Decennial Census collected statistical data on the Hispanic population by directly asking each respondent to identify their Hispanic origin. Use of this self-identifier approach,

however, may be impractical or impossible in other surveys, requiring another method of detecting the Hispanic population. One such method involves matching the respondents' surnames against a list of common Hispanic surnames. Further research into this method led to the Passel-Word Spanish surname list, developed by Jeffrey Passel and David Word for the 1980 Decennial Census. The continued growth of the Hispanic population in the United States during the 1980's and 1990's prompted an evaluation of the Passel-Word list, to see if it retained its effectiveness in identifying the Hispanic population. This report evaluates the performance of the Passel-Word Spanish surname list in determining the Hispanic origin of respondents found in a sample of the 1990 Decennial Census Post Enumeration Survey data.

HISTORY

The United States Bureau of the Census has used Spanish surname lists as a method of identifying the Hispanic population for more than 40 years. In 1950, the first Spanish surname list helped indicate the Hispanic population found in Arizona, California, Colorado, New Mexico, and Texas. New Spanish surname lists developed whenever additional significant Spanish surname data became available. In his 1975 United States Bureau of the Census technical paper, entitled Comparison of Persons of Spanish Surname and Persons of Spanish Origin in the United States, Edward Fernandez analyzed the effectiveness of the 1970 Census Spanish surname list in identifying the Hispanic population in the five aforementioned Southwestern states. He concluded that while the list was an adequate measure of Spanish origin for these five states, it fared badly in identifying Hispanics in the remainder of the country. Later research by Jeffrey Passel and David Word, as documented in their paper, Constructing the List of Spanish Surnames for the 1980 Census: An Application of Bayes' Theorem, concluded that the list's difficulties stemmed more from the surnames on the list than from the concept of surname lists.

Consequently, for the 1980 Decennial Census, Passel and Word developed a Spanish surname list. They started with the premise that Spanish surnames and the United States' Hispanic population share a similar geographic distribution. Passel and Word divided the country into 858 mutually exclusive geographic areas, creating both highly Spanish and highly non-Spanish areas. For each area, they assigned a numerical geographic weight, a logarithmic comparison of the area's concentration of the Hispanic population as compared to the concentration of the Hispanic population in the country. The 1977 Federal Income Tax Returns provided a large national list of surnames with geographic identifiers. A statistical function based on Bayes' Theorem and the multinomial distribution used the geographic weights to assign values of Hispanic similarity to each unique surname from the tax returns. From the initial 1.4 million distinct surnames, the statistical function assigned 16,514 surnames to the preliminary Spanish surname list.

Non-Spanish surnames had a better chance of making the preliminary list if they occurred infrequently within the Federal Tax Returns. Additional tests were performed on the preliminary list to weed out these surnames. Surnames more commonly associated with other ethnic groups were eliminated, as well as surnames with letters not found in the Spanish alphabet. Those infrequently occurring surnames commonly found among the small number of Puerto Rico returns remained on the list. Surnames also stayed on the list if they appeared on the 1970 Census Spanish surname list or an extended surname list developed after consultation with

linguistic experts evaluating the 1970 Census list. The resulting 12,497 Hispanic surnames became the Passel-Word (PW) Spanish surname list.

To compare the effectiveness of their list against the 1970 Census Spanish surname list, Passel and Word utilized the March 1976 Current Population Survey (CPS) results. They ran the surnames from the CPS records through each surname list and recorded whether or not the surname was found on either list. They then compared these matching results against the Hispanic identifier in the CPS records. Table 1 presents the results of the CPS test.

Table 1. Comparison of Surname Lists - March 1976 CPS Test

	<u>Overall Results</u>	<u>Southwestern States</u>	<u>Remainder of United States</u>
Percent of Respondents With Hispanic Surnames Detected By Surname List Who Reported a Non-Hispanic Origin:			
1970 Census List	30.4%	15.3%	46.0%
Passel-Word List	15.0%	10.5%	21.4%
Percent of Respondents Who Reported a Hispanic Origin That Have Surnames Not Found On Surname List:			
1970 Census List	26.3%	21.6%	33.0%
Passel-Word List	20.7%	16.1%	27.1%

The PW list's lower percentages indicate that it performed better than the 1970 Census list in both the five Southwestern states and the remainder of the country in two important measures. First, of those CPS records with surnames included on the PW list and 1970 Census list, roughly 15.0 percent and 30.4 percent, respectively, came from Non-Hispanic respondents. Fewer matched Non-Hispanic records result from an increase of strong Hispanic surnames and/or a decrease of weak Hispanic surnames on a Spanish surname list. Second, the PW list did not identify 20.7 percent of the surnames belonging to self-reported Hispanic CPS respondents as Hispanic, compared to 26.3 percent for the 1970 Census list. Fewer missed Hispanic records result from an increase of Hispanic surnames on a Spanish surname list. These results led to the use of the PW list as the Spanish surname list for the 1980 Decennial Census.

RESEARCH BACKGROUND

An effective evaluation of the PW Spanish surname list requires a large number of records evenly distributed across the United States which contains, at minimum, respondents' surnames and their Hispanic origins. Even when used internally, the basic Decennial Census record does not include the respondent's surname, prompting the need for a file that does include surnames.

The 1990 Spanish Origin Research (SOR) file meets these requirements. The SOR file contains unedited and unallocated data taken from a sample of the 1990 Decennial Census questionnaires. The SOR sample includes all housing units and group quarters in or near the Census-defined blocks sampled for the Post Enumeration Survey. The Post Enumeration Survey was conducted by the United States Bureau of the Census to provide estimates of the 1990 Decennial Census undercount.

Each SOR record contains the geographic and demographic data for each individual listed on the sampled questionnaire. Geographic data found on the record includes codes for the state, county, and block where the individual lives. The SOR file covers data collected for all 50 states and the District of Columbia. Unedited and unallocated demographic data found on the record covers the population questions asked on all 1990 Decennial Census questionnaires. These demographic data include sex, age, race, Hispanic origin, marital status, and relationship to householder. Unedited and unallocated data means that none of the Census edit and allocation procedures have been performed on the data. The SOR data comes directly from the individual's answers on the questionnaire, including blank fields for those questions the individual did not answer.

Before proceeding with the PW Spanish surname list evaluation, a name standardization program was developed to process each individual's full name as found in the SOR file. The surname matching program used in the evaluation required exact matches, making the standardization of surnames to the PW list format important to improve matching efficiency. The standardization program first identified each element, defined as a character string delimited by blanks, recorded in an individual's first name, middle initial, and surname fields. It determined the base name for the first name and surname fields, labelling any additional elements as prefixes and suffixes. For records with compound surnames, it marked the first surname as the base name and the additional surnames as suffixes. It then formatted each base name, if needed, to match the PW list surname format, possibly changing the names recorded in the other fields. Formatting procedures included eliminating prefix and suffix titles, compressing common surname prefixes into the base surname, removing middle names from the first name and surname fields, and recognizing entries such as "ADULT MALE" as non-names.

Only SOR records with a standardized surname and a valid Hispanic origin code qualified for the PW Spanish surname list evaluation. Blank or unknown values in either field disqualified the record. The SOR file initially contained 7,154,390 records, but only 5,609,592 records (78.4 percent) met the requirements for the evaluation. The reason for eliminating most SOR records was respondents failing to include their surname and/or Hispanic origin. Additionally, the name standardization process eliminated 29,681 records (0.4 percent) and changed either the first name or surname field on an additional 198,104 records (2.8 percent). Within this report, the term "matchable SOR records" will define those SOR records with standardized surname and valid Hispanic origin codes.

The evaluation of the PW Spanish surname list began by the collection of all matchable SOR records, complete with their geographic and demographic data, into a single database. A surname matching program then read each standardized surname in this database and looked for an exact match within the PW list. This program concluded by adding an additional record match field to the database, storing the results of whether the PW list contained an exact match or not. Separate

programs created frequency tables from the SOR records that compared the PW list match results to the self-reported Hispanic origin answers. These tables served as the source of the statistics used to judge the effectiveness of the PW list. Geographic and demographic variables found on the SOR records served as partitions for the database when creating the frequency tables. Only those matchable SOR records with a valid non-blank value for the partitioning variable were used to create that variable's frequency tables.

Two important statistics judge the effectiveness of the PW Spanish surname list. The surname commission (SCOM) rate defines the percentage of people whose surnames appeared on the PW list that reported a Non-Hispanic origin. The lower the SCOM rate, the more reliable the surnames on the PW list are at detecting the Hispanic population. The surname omission (SOM) rate defines the percentage of people who reported a Hispanic origin that had a surname not appearing on the PW list. The lower the SOM rate, the greater proportion of the Hispanic population the PW list finds. For example, using the percentages reported in Table 1, the PW list had a 15.0 percent SCOM rate and a 20.7 percent SOM rate for the March 1976 CPS data. Surname omission rates can be calculated for the specific Hispanic origin categories as well. For example, the Mexican SOM rate defines the percentage of people who reported a Mexican origin but whose surnames did not appear on the PW Spanish surname list.

MATCHING RESULTS

Overall Results

The PW Spanish surname list performed better on the 1990 SOR data than it did on the March 1976 CPS data. The PW list has a 10.02 percent surname commission rate and a 20.55 percent surname omission rate for the 1990 SOR data. The standard errors for these estimates are 0.04 percent and 0.05 percent, respectively. Although the SOR data's SOM rate compares with favor to the CPS data's SOM rate (20.73 percent), the SOR data's SCOM rate is a one-third reduction over the CPS data's SCOM rate (15.02 percent). There was no mention of either standard errors or the number of respondents processed during the March 1976 CPS test in Passel and Word's findings.

A larger, more tightly clustered Hispanic population existed in 1990 than in 1976, partially explaining the PW list's improved performance on the 1990 SOR data. Several survey-related differences exist between the 1976 CPS and 1990 SOR data that also may help explain the differences in their surname error rates: survey methods, sampling population, and sampling design.

The 1976 CPS used personal and telephone interviews to collect its data, while the 1990 Census relied mostly on mailback questionnaires. Enumerators conducting personal and telephone interviews have the advantage of eliminating any confusion surrounding the questions asked, but they can possibly, even unknowingly, influence the respondents' answers. Some respondents may answer questions incorrectly if they feel pressured by an enumerator. Other respondents may give answers they feel the enumerator, as a member of the Census Bureau and the federal government, wishes or wishes not to hear. A questionnaire allows the respondents time to answer the questions, preventing possible enumerator bias, but it might return less information than an

interview. Respondents might not answer questions if they either do not understand it or simply refuse to answer it.

While the 1990 Decennial Census SOR data set contains a cross-section of the country's population, the March 1976 CPS data set used to test the PW list contained only respondents 14 years of age or older. Based on the 1990 SOR data, the SCOM rate for children 13 years and younger was 8.81 percent. Without this group, the SOR data's overall SCOM rate would have been 10.59 percent. At this age, children still have their parents' surname, which more reliably reflects their Hispanic origin (if any) than any other surname they may acquire in their lives. Children have a very small probability of changing their surname legally, such as through marriage or a court-approved name change. Thus, children have low surname error rates, partially causing the SOR data's lower SCOM rate.

Although the number of individual records was not reported, the March 1976 Current Population Survey interviewed 50,000 households, with a double sampling of Hispanic households. Estimating the number of individuals in the March 1976 CPS data file and comparing it to the 1990 SOR file, the SOR file contained approximately 50 times more records than the CPS file. The more surname records, the more reliable the surname statistics; although taking a large sample did not guarantee lower SCOM rates, it presented a better estimate of the true SCOM rate.

Hispanic Origin

Of the individuals listed in the 5,609,592 matchable SOR records, 597,533 individuals (10.7 percent) reported a specific Hispanic origin. Table 2 compares the 1990 SOR data's PW Spanish surname list matching results by Hispanic origin.

Table 2. PW Spanish Surname List Performance - Hispanic Origin

	Number and Percent of Hispanic SOR Records	SURNAME OMISSION	
		<u>Rate</u>	<u>Std Error</u>
Mexicans	359,293 (60.1%)	16.10%	0.06%
Puerto Ricans	95,298 (15.9%)	18.84%	0.13%
Cubans	23,094 (3.9%)	26.12%	0.29%
Other Hispanic	119,848 (20.1%)	34.18%	0.14%

The Mexican population has the smallest specific Hispanic SOM rate at 16.10 percent. Mexicans form the single largest Hispanic population in the United States, comprising 60.1 percent of the self-reported Hispanics found in the 1990 SOR file. They reside mostly in states adjacent to the Mexican border, such as Arizona, California, New Mexico, and Texas. Approximately 69.8 percent of the Mexicans sampled in the SOR file live in either California or Texas. When creating their Spanish surname list, Passel and Word wanted to include only those Spanish surnames that have a geographical distribution similar to the distribution of the Hispanic population in the United States. Because of the heavy concentration of Mexicans in the

Southwest region, common Spanish surnames from this region made the list, leading to low Mexican SOM rates.

Puerto Ricans and Cubans also have similar concentrated areas in the United States, although not as large as the Mexican population. Puerto Ricans form the second largest Hispanic population in the country, comprising 15.9 percent of the self-reported Hispanics found in the 1990 SOR file. They live mostly in the Northeast, particularly in Illinois, New Jersey, and New York. Their SOM rate is 18.84 percent, 1.71 percentage points smaller than the overall SOM rate, but 2.74 percentage points greater than the Mexican SOM rate. Cubans make up only 3.9 percent of the self-reported Hispanics, their largest concentration residing in Florida. They have a 26.12 percent SOM rate, 5.57 percentage points greater than the overall SOM rate.

The Other Hispanic group consists of self-reported Hispanic respondents who do not consider themselves Mexicans, Puerto Ricans, or Cubans. They have the largest specific Hispanic SOM rate at 34.18 percent. Two reasons may explain this high SOM rate. First, any unique surnames that fall within this group had a small probability of ending up on the PW Spanish surname list. Respondents in the Other Hispanic category may have origins based in Central or South American countries or in Spain. They may have a broader array of surnames than the other three Hispanic categories. However, unless concentrated in a heavily Hispanic-populated state, they were not large enough to influence the selection of surnames on the PW list at the time of its creation. Second, respondents who had difficulties answering the Hispanic question during the 1990 Decennial Census may have selected this group. The latter situation more likely occurred than the former situation. Previous Census research of response variabilities during reinterview studies, as noted by Passel and Word in their 1980 Spanish surname list research, showed that up to 30 percent of respondents that initially reported the Other Hispanic category as their Hispanic origin will report a Non-Hispanic origin when reinterviewed. Research into the written Hispanic origin responses that accompanied the Other Hispanic records could better explain the reasons for the high Other Hispanic SOM rate.

A strict comparison between the specific Hispanic SOM rates of the March 1976 CPS data and the 1990 Decennial Census SOR data would not be statistically viable. The CPS included a Central or South American category as one of its Hispanic origin choices. Respondents with Central or South American origins would have chosen the Other Hispanic category on the 1990 Decennial Census questionnaires. The mixture of these respondents into the SOR Hispanic origin categories prevented a fair comparison with similar categories from the CPS test.

Table 3. PW Spanish Surname List Performance - States

<u>State</u>	<u>SURNAME COMMISSION</u>		<u>SURNAME OMISSION</u>	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Alabama	44.26%	1.86%	63.05%	1.47%
Alaska	41.51%	2.39%	54.41%	2.14%
Arizona	10.85%	0.23%	22.50%	0.29%
Arkansas	35.67%	2.18%	48.68%	2.03%
California	8.43%	0.07%	17.74%	0.09%

Colorado	12.21%	0.34%	21.67%	0.41%
Connecticut	13.07%	0.50%	20.16%	0.57%
Delaware	20.65%	2.02%	33.95%	2.15%
District of Columbia	15.40%	1.81%	33.79%	2.10%
Florida	10.15%	0.18%	26.06%	0.24%
Georgia	21.14%	0.92%	39.82%	0.96%
Hawaii	61.39%	0.79%	59.54%	0.82%
Idaho	18.74%	0.91%	32.16%	1.00%
Illinois	5.87%	0.13%	15.99%	0.19%
Indiana	23.98%	1.50%	38.80%	1.54%
Iowa	30.46%	2.01%	50.48%	1.85%
Kansas	12.96%	0.68%	28.64%	0.82%
Kentucky	26.35%	1.04%	45.47%	1.02%
Louisiana	40.73%	0.89%	42.86%	0.88%
Maine	57.14%	3.07%	70.94%	2.32%
Maryland	30.52%	1.38%	45.84%	1.32%
Massachusetts	35.83%	0.79%	34.40%	0.79%
Michigan	20.05%	0.88%	37.57%	0.94%
Minnesota	29.62%	1.57%	44.33%	1.52%
Mississippi	66.80%	2.09%	69.06%	1.98%
Missouri	41.15%	1.76%	54.73%	1.56%
Montana	41.69%	2.32%	48.02%	2.22%
Nebraska	23.05%	1.66%	31.96%	1.73%
Nevada	17.55%	0.68%	29.38%	0.75%
New Hampshire	53.51%	2.59%	59.34%	2.39%
New Jersey	9.08%	0.25%	21.23%	0.33%
New Mexico	6.84%	0.17%	15.90%	0.24%
New York	8.38%	0.11%	22.75%	0.15%
North Carolina	21.48%	0.55%	36.55%	0.58%
North Dakota	60.87%	10.18%	75.00%	7.22%
Ohio	25.20%	1.04%	37.62%	1.06%
Oklahoma	21.23%	0.89%	37.93%	0.93%
Oregon	19.58%	0.76%	33.26%	0.82%
Pennsylvania	13.67%	0.43%	23.13%	0.50%
Rhode Island	51.60%	1.66%	40.70%	1.80%
South Carolina	43.28%	2.45%	54.15%	2.22%
South Dakota	58.06%	8.86%	50.00%	9.81%
Tennessee	37.45%	1.82%	58.20%	1.52%

Texas	4.22%	0.06%	11.99%	0.09%
Utah	24.75%	1.04%	40.05%	1.06%
Vermont	66.67%	4.15%	63.87%	4.40%
Virginia	31.47%	0.91%	44.87%	0.87%
Washington	22.34%	0.52%	30.67%	0.55%
West Virginia	60.21%	2.88%	55.43%	3.09%
Wisconsin	17.42%	1.05%	24.67%	1.14%
Wyoming	17.64%	1.22%	33.61%	1.35%

States

Table 3 compares the 1990 SOR data's PW Spanish surname list matching results for all matchable SOR records by state, including the District of Columbia.

A fair inverse correlation exists between the Hispanic concentration within a state and that state's SCOM rate. The Pearson's correlation coefficient between the two variables stands at -0.58, statistically significant evidence that an inverse trend exists between them. The larger the proportion of Hispanics within the state, the smaller the SCOM rate; the smaller the proportion, the larger the SCOM rate. This correlation suggests the strength of established Hispanic communities. These communities become strong Hispanic cultural bases, attracting immigrant Hispanics and promoting Hispanic families. Hispanics who live in such communities are easier to identify with surname lists than Hispanics who live elsewhere. Established Hispanic communities exist more within states of high Hispanic concentration than in states of low Hispanic concentration, causing the associated SCOM rates. An exception to this theory comes when a state has a strong presence of a Non-Hispanic population with Hispanic surnames, such as Native Hawaiians and Filipinos.

A stronger inverse correlation also exists between the Hispanic concentration within a state and that state's SOM rate. The Pearson's correlation coefficient between the two variables stands at -0.66, statistically significant evidence that an inverse trend exists between them. Established communities usually cater to a specific Hispanic origin, however, making individuals of other Hispanic origins possibly more difficult to detect. For example, Florida serves as the home for several established Cuban communities. The state's Cuban SOM rate is 22.0 percent, while the combined SOM rate for the remaining Hispanic origins is 29.9 percent. The PW Spanish surname list works well at identifying specific Hispanic origins within their states of high concentration, but not as well outside those states.

Geographic Regions

Table 4 compares the 1990 SOR data's PW Spanish surname list matching results for all matchable SOR records by geographic region. The Mexican Southwest group includes Arizona, California, Colorado, New Mexico, and Texas. The Puerto Rican Northeast group includes Connecticut, Illinois, New Jersey, New York, and Pennsylvania.

Table 4. PW Spanish Surname List Performance - Geographic Region

<u>Geographic Region</u>	<u>SURNAME COMMISSION</u>		<u>SURNAME OMISSION</u>	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Mexican Southwest	7.11%	0.05%	16.05%	0.06%
Puerto Rican Northeast	8.26%	0.08%	20.86%	0.11%
Florida	10.15%	0.18%	26.06%	0.24%
Hawaii	61.39%	0.79%	59.54%	0.82%
Rest of United States	26.74%	0.19%	38.57%	0.19%

Among the defined geographic regions, the PW Spanish surname list performs best in the Mexican Southwest region. The region's SCOM and SOM rates are 7.11 percent and 16.05 percent, respectively. The large number of Mexicans living in the region explain the low surname error rates. Approximately 80.8 percent of the Mexican respondents sampled in the 1990 SOR file live in this region. As noted in The Hispanic Population of the United States, a book written by Frank Bean and Marta Tienda that analyzed the 1980 Decennial Census Hispanic demographic data, the United States victory in the Mexican-American war resulted in the creation of the five states that form the Mexican Southwest region. The Mexican communities established in this region became part of the country, and a large Mexican population continues to live there today. Numerous surnames from this region appear on the PW Spanish surname list because of these large concentrated Mexican communities. The Mexican SOM rate for this region is 13.91 percent, while the Mexican SOM rate for the remainder of the country is 25.35 percent.

Another region that contributed a number of surnames to the PW Spanish surname list is the Puerto Rican Northeast region. Approximately 79.4 percent of the Puerto Rican respondents sampled in the 1990 SOR file live in this region. Bean and Tienda noted that the wave of Puerto Ricans that entered the United States during the 1950's settled in the Northeast, providing their industries with cheap labor. The communities the Puerto Ricans established in this region led to its high Hispanic concentration, which aided placing its common surnames on the PW list. The Puerto Rican Northeast's 8.26 percent SCOM rate falls below the overall average SCOM rate, but its 20.86 percent SOM rate is 0.31 percentage points above the overall average SOM rate. The Puerto Rican SOM rate for this region is 16.14 percent, while the Puerto Rican SOM rate for the remainder of the country is 29.26 percent.

Unlike Puerto Ricans, the first waves of Cubans that immigrated to the United States in the 1950's resided primarily in Florida, the state nearest their homeland. According to Bean and Tienda, they quickly established cultural communities within the state, providing a home for future Cuban immigrants. Approximately 69.0 percent of the Cuban population sampled by the SOR files lives in Florida. As an effect of these communities, most unique Hispanic surnames concentrated in this state ended up on the PW Spanish surname list. Florida's Cuban SOM rate stands at 22.05 percent, 4.07 percentage points smaller than the overall Cuban SOM rate, while the Cuban SOM rate for the remainder of the country rests at 35.18 percent. Despite the state's low Cuban SOM rate, its overall SCOM and SOM rates are 10.15 percent and 26.06 percent, respectively, both rates greater than their national averages.

Hawaii presents a special problem for the PW Spanish surname list. While 10.1 percent of the state's population sampled in the SOR file considered themselves having a Hispanic origin, Hawaii's SCOM and SOM rates are 61.39 percent and 59.54 percent, respectively. Hawaii's geographic location partially explains its high SCOM rate. Located within the Pacific Ocean, Hawaii serves as a home for many of the Asian and Pacific Island races. Two races in particular, Native Hawaiians and Filipinos, have surnames that appear on the PW list. Separated from the other races, Native Hawaiians and Filipinos constitute 30.5 percent of the sampled Hawaiian respondents. More importantly, they form 53.4 percent of the sampled Hawaiian respondents with surnames found on the PW list. Although not considered Hispanics, the Native Hawaiian and Filipino cultures incorporate many Hispanic surnames, resulting in a combined 82.65 percent SCOM rate within Hawaii. Without the respondents of these two races, Hawaii's SCOM rate drops to 39.93 percent, similar to states with very low Hispanic population proportions.

Hawaii's high SOM rate, however, could not be explained by the numerous races populating the state. Hawaii's overall 59.54 percent SOM rate is approximately 2.9 times greater than the national SOM rate. Among the individual race categories in the state with more than 10 self-reported Hispanics, the SOM rates range from 40.00 percent (for American Indians) to 85.11 percent (for Blacks). Hispanics living in this multicultural state appear to take surnames from other sources than their ethnicity. One possible theory is that marriages between members of different ethnic groups are more acceptable within the multicultural state, making it more difficult to detect Hispanic origin via a surname list. Based on these findings, it appears doubtful that the PW list, or any other Spanish surname list, would work in identifying Hawaii's true Hispanic population.

The remaining 38 states and the District of Columbia combine for a 26.74 percent SCOM rate and a 38.57 percent SOM rate. These surname error rates are 16.69 percentage points and 18.02 percentage points, respectively, higher than the national rates. The low concentration of Hispanics in this region helps decipher its high surname error rates. Within their combined SOR sample, only 2.2 percent of their respondents claim a Hispanic origin. Fewer established Hispanic communities exist in this region than any of the other defined regions, leading to a stronger blend of Hispanics within the Non-Hispanic communities. Common Spanish surnames found only in this region, if any, had a small probability of making the PW Spanish surname list in its development. The Hispanic respondents in these states, away from their focal points of cultural heritage and traditions, may adapt more easily to American culture. Side effects of this adaptation include marrying a Non-Hispanic mate or Americanizing their surnames, making it harder to identify their origin correctly and pushing up the surname error rates.

Relationships

Table 5 compares the 1990 SOR data's PW Spanish surname list matching results for all matchable SOR records where respondents recorded their relationship to the householder. Relatives not usually found in the household, like the householder's parents, siblings, and grandchildren, are grouped together in the "Other Relatives" category. Non-relatives are all persons living within the household not related to the householder, including roommates, housemates, roomers, foster children, and unmarried partners. Group quarters include people living in orphanages, nursing homes, prisons, dormitories, and boarding houses.

Table 5. PW Spanish Surname List Performance - Relationship to Householder

<u>Relationship</u>	<u>SURNAME COMMISSION</u>		<u>SURNAME OMISSION</u>	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Householder	9.66%	0.08%	19.44%	0.10%
Husband/Wife	17.07%	0.13%	26.17%	0.15%
Son/Daughter	8.45%	0.06%	18.97%	0.08%
Stepchildren	10.83%	0.34%	23.40%	0.42%
Other Relatives	6.97%	0.11%	17.95%	0.16%
Non-Relatives	8.21%	0.16%	22.63%	0.23%
Group Quarters	13.57%	0.33%	27.06%	0.39%

The most noteworthy item of interest from Table 5 concerns the PW Spanish surname list differences when it comes to identifying the Hispanic origins of householders and their spouses. Householders have SCOM and SOM rates of 9.66 percent and 19.44 percent, respectively, while their spouses have SCOM and SOM rates of 17.07 percent and 26.17 percent, respectively. The higher spouse miss rates most likely result from marital surname changes. Most married couples complete the questionnaires by signifying the male as the head of the household and the female as his spouse. Males make up 67.1 percent of the householders in the SOR file, while females make up 92.2 percent of the spouses. When either a Non-Hispanic wife takes her husband's Hispanic surname or a Hispanic wife takes her husband's Non-Hispanic surname, the PW Spanish surname list will incorrectly identify the spouse's origin. No known modifications to the surname list will correct this problem, although an adjoining list of strong Hispanic first names, like "Jesus" and "Blanca", might lower the SOM rate. Similar surname problems would occur in those marriages where the husband takes the wife's surname.

Sex

Table 6 compares the 1990 SOR data's PW Spanish surname list matching results for all matchable SOR records where respondents recorded their gender.

Table 6. PW Spanish Surname List Performance - Sex

<u>Sex</u>	<u>SURNAME COMMISSION</u>		<u>SURNAME OMISSION</u>	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Male	7.71%	0.05%	18.25%	0.07%
Female	12.41%	0.07%	22.88%	0.08%

The PW Spanish surname list correctly identifies a larger share of males than females within the SOR file. Non-Hispanic males and females comprise 7.71 percent and 12.41 percent, respectively, of the Spanish surname matches made. Among the Hispanic respondents, the surname list failed to identify 18.25 percent and 22.88 percent, respectively, of the Hispanic males and females. The higher female miss rates most likely result from the tradition that wives

take their husbands' surnames after they marry. The PW list will misclassify a married female respondent if either a Non-Hispanic wife takes her husband's Spanish surname or a Hispanic wife takes her husband's non-Spanish surname. Similar differences between both the SCOM and SOM rates suggest both cases occur equally as often.

The effects of marriage on identifying female Hispanics by surnames can be seen when dividing each sex into marital status classifications. Among the never married males and females of all ages listed in the SOR file, never married females have a slightly higher SCOM rate (8.62 percent) than never married males (8.18 percent). Never married males and females have similar SOM rates (19.95 percent and 20.02 percent, respectively). A comparison of now married males and females in the same file finds that married females have a significantly greater SCOM rate (17.63 percent) and SOM rate (26.56 percent) than married men (7.17 percent and 15.42 percent, respectively). These comparisons show that surnames taken after marriage make it more difficult to properly identify female respondents. Acquiring and matching the maiden names of married females should improve the performance of the surname matching process. The research on Hispanic first names may also help to overcome this weakness.

Race

Table 7 compares the 1990 SOR data's PW Spanish surname list matching results for all matchable SOR records where respondents recorded their race. American Indians, Eskimos, and Aleutians form the AIEA category, while the various Asian and Pacific Island races form the API category.

Table 7. PW Spanish Surname List Performance - Race

<u>Race</u>	SURNAME COMMISSION		SURNAME OMISSION	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Whites	14.38%	0.07%	23.05%	0.08%
Blacks	49.33%	0.59%	66.05%	0.46%
AIEA	33.22%	0.64%	29.22%	0.63%
API	48.16%	0.34%	32.49%	0.37%
Other Races	1.20%	0.03%	16.36%	0.08%

Those SOR respondents that consider Hispanic a separate race answered the race question by checking the "Other Races" category. While the "Other Races" category covers 4.0 percent of all SOR respondents, it contains 39.7 percent of all Hispanic respondents that answered the race question. The PW Spanish surname list proves effective in detecting the large number of Hispanics in this category, resulting in the category's low surname error rates. The SCOM and SOM rates for races not separately listed on the Census questionnaire are 1.20 percent and 16.36 percent, respectively. The sizable proportion of Hispanic respondents found in the "Other Races" category and its surname error rates verify that many Hispanic SOR respondents considered their ethnicity as their race.

Whites dominate the SOR records, comprising 76.5 percent of all respondents and 53.6 percent of the Hispanic respondents. The SCOM and SOM rates for Whites are 14.38 percent and 23.05 percent, respectively. These rates would be smaller if the SOR respondents could not classify their Hispanic ethnicity as their race. Black Hispanics prove most difficult for the PW Spanish surname list to identify. The small percentage of Black Hispanics in the SOR file, as compared to Black Non-Hispanics, help explain the 49.33 percent SCOM rate and 66.05 percent SOM rate. While Blacks define 16.0 percent of the SOR respondents, they only define 2.2 percent of the Hispanic respondents. The high surname error rates suggest that Black Hispanics integrate themselves better into the American culture than Hispanic members of other races.

The SCOM and SOM rates for the AIEA category are 33.22 percent and 29.22 percent, respectively. With American Indians comprising 97.6 percent of the AIEA data, mixed American Indian and Hispanic cultures located in the Southwestern states may explain these surname error rates. Being roughly two similar cultures, mixed communities of American Indians and Mexicans (or other Hispanic races) could lead to Non-Hispanic American Indians with Hispanic surnames and Hispanic American Indians with Non-Hispanic surnames, causing increases in both AIEA surname error rates.

The API group has a 48.16 percent SCOM rate and 32.49 percent SOM rate. A comparison between the individual API races and the "Other API" category shows some unusual results. The SCOM rates for the individual API races range from 43.55 percent for Vietnamese to 91.06 percent for Filipinos, while their SOM rates range from 40.31 percent for Guamanians to 88.11 percent for Chinese. The "Other API" category, which covers all API races not listed separately on the 1990 Decennial Census questionnaire, has a 3.34 percent SCOM rate and a 21.20 percent SOM rate. Without this category, the individual API races have a combined 88.04 percent SCOM rate and 66.76 percent SOM rate.

Two surname patterns emerge from the API group. First, the Native Hawaiian and Filipino races share many of the surnames found on the PW Spanish surname list. Although not considered Hispanic races, the Native Hawaiian and Filipino cultures have been influenced by the Spanish within their pasts. The two races, which make up 24.0 percent of the SOR file's API respondents, have a combined 89.94 percent SCOM rate and 65.36 percent SOM rate. Deletion of surnames that are significantly more popular among Native Hawaiians and Filipinos than Hispanics will lower both the API and overall surname error rates. Second, the "Other API" category may contain a number of incorrect race responses. This catch-all category, which makes up 17.8 percent of the SOR file's API respondents, includes 75.2 percent of the self-reported API Hispanics. On 1990 Decennial Census questionnaires, the write-in box for the "Other API" response is also the write-in box for the "Other Race" response. Some Hispanics may have filled in the "Other API" circle accidentally when they recorded their race as Hispanic. It is also possible that a Hispanic-influenced API race has become established within the United States, but the former case seems much more likely. An examination of Hispanic API surnames and/or the write-in "Other API" responses would benefit the future use of the PW Spanish surname list.

Marital Status

Table 8 compares the 1990 SOR data's PW Spanish surname list matching results for all matchable SOR records where respondents recorded their marital status. The "Were Married" group includes divorced, widowed, and separated respondents.

Table 8. PW Spanish Surname List Performance - Marital Status

<u>Marital Status</u>	<u>SURNAME COMMISSION</u>		<u>SURNAME OMISSION</u>	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Now Married	12.18%	0.08%	20.83%	0.09%
Were Married	12.19%	0.14%	23.16%	0.17%
Never Married	8.39%	0.05%	20.01%	0.07%

Never married respondents have the lowest SCOM rate at 8.39 percent. Now married respondents have a 12.18 percent SCOM rate, while those respondents that were married have a 12.19 percent SCOM rate. Non-Hispanic spouses who take the surname of their Hispanic partner upon marriage prompt the PW Spanish surname list to classify them incorrectly as Hispanic, causing the now and were married respondents' higher surname commission rates.

Never married respondents also emerge with the significantly lowest SOM rate at 20.01 percent. Now married respondents show a 20.83 percent SOM rate, while those respondents that were married show the highest SOM rate at 23.16 percent. Hispanic spouses taking the surname of their Non-Hispanic partner again affects the SOM rates among the marital statuses the most, causing the lower SOM rate among the never married respondents. A look among the surname omission rates for the specific Hispanic populations, however, shows no clear trend among the three marital statuses. Never married respondents have the smallest Other Hispanic SOM rate at 32.98 percent. Now married respondents have the smallest Mexican SOM rate at 15.69 percent, while respondents that were married have both the smallest Puerto Rican (18.25 percent) and Cuban (24.33 percent) SOM rates. The percentage breakdown of Hispanics within each status also affects their overall SOM rates.

Age

Table 9 compares the basic PW Spanish surname list matching results for all matchable SOR records where respondents recorded their age.

Table 9. PW Spanish Surname List Performance - Age Group

<u>Age Group (Years)</u>	<u>SURNAME COMMISSION</u>		<u>SURNAME OMISSION</u>	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Children (0-18)	8.47%	0.06%	19.85%	0.09%
Young Adults (19-30)	10.20%	0.08%	21.03%	0.11%
Adults (31-44)	11.80%	0.10%	21.59%	0.12%
Mature Adults (45-64)	11.06%	0.12%	20.10%	0.15%
Elderly Adults (65+)	12.49%	0.21%	20.53%	0.25%

A similar pattern exists between SCOM and SOM rates when broken down into the defined age groups. Children have the lowest SCOM (8.47 percent) and SOM (19.85 percent) rate among all the age groups. Both surname error rates increase as the age group gets older - with the exception of mature adults. An unexplained decrease in both the SCOM and SOM rate occurs from adults to mature adults. Despite having similar patterns, no trend appears to exist among the defined age groups as measured by the SCOM and SOM rates. The surname error rates for mature and elderly adults break the pattern established by the younger age groups.

Several factors might explain the surname error rate pattern established by the defined age groups. First, the surname handed from parents to children is the best surname identifier of Hispanic origin. Children have the least chance of changing their legal surnames, prompting the lowest SCOM and SOM rates. Second, the initial increases in the SCOM and SOM rates from children to adults come mostly from women that take new surnames upon marriage. Non-Hispanic women taking their Hispanic husband's surname increases the SCOM rate, while Hispanic women taking their Non-Hispanic husband's surname increases the SOM rate. Third, the elderly adults age category contains more females than males due to men's shorter life expectancies. Females comprise 59.2 percent of the SOR file respondents of age 65 years or older. Within this age group, female elderly adults have a higher SCOM rate (13.69 percent) and a higher SOM rate (23.14 percent) than male elderly adults (10.83 percent and 16.74 percent, respectively). Having higher surname error rates than males, due to marital surname changes, the larger proportion of elderly adult females translates into an increase in the age group's surname error rates.

EVALUATION

One way of measuring the effectiveness of the PW Spanish surname list is to compare the attributes of the true Hispanic population to the population determined by the PW list as Hispanic. If both populations' attributes are similar, then it can be said that the PW list identifies a population that contains Hispanic attributes. Table 10 compares the three standard deviation ranges of the geographic and demographic point estimate compositions of two SOR record lists. The first list contains those SOR records that claim a Hispanic origin, while the second list contains those SOR records where the respondents' surnames fall on the PW list.

The PW Spanish surname list detects a population similar to the Hispanic population when investigating the sex, age, and householder relationship demographics. The overlapping of three standard deviation ranges between the Hispanic origin and matched Spanish surname lists denotes a similarity between attributes, the size of the overlap an estimated measure of the strength of similarity. The greatest disparities between the two groups came in geographical location and the race demographic.

A better understanding at where the sampled Hispanic population and the Hispanic population as determined by the PW Spanish surname list differ comes from examining where the two populations do and do not overlap. Table 11 examines the geographic and demographic percentage compositions for various subsets of the SOR file. The first subset contains SOR respondents who reported a Hispanic origin and whose surnames were on the PW list. The second subset contains respondents with Hispanic origins but unmatched surnames, while the

third subset contains respondents with matched surnames but Non-Hispanic origins. Respondents in the second subset contributed to the PW list's surname omission rates, while respondents in the third subset contributed to its surname commission rates.

**Table 10. Comparing Attributes of Hispanic Origin
Against Passel-Word Surname List**

<u>Variable</u>	HISPANIC ORIGIN			PW SURNAME LIST		
	<u>Minimum</u>	<u>Estimate</u>	<u>Maximum</u>	<u>Minimum</u>	<u>Estimate</u>	<u>Maximum</u>
GEOGRAPHY:						
Mexican Southwest	58.22%	58.41%	58.60%	59.59%	59.79%	60.00%
Puerto Rican Northeast	24.12%	24.28%	24.45%	23.55%	23.72%	23.90%
Florida	5.38%	5.47%	5.56%	5.01%	5.10%	5.19%
Hawaii	0.57%	0.60%	0.63%	0.68%	0.71%	0.75%
Rest of United States	11.12%	11.24%	11.36%	10.55%	10.67%	10.80%
RELATIONSHIP:						
Householder	27.56%	27.74%	27.91%	27.80%	27.99%	28.18%
Husband/Wife	15.12%	15.26%	15.40%	15.22%	15.37%	15.53%
Son/Daughter	37.06%	37.25%	37.44%	37.10%	37.31%	37.51%
Stepchildren	1.68%	1.73%	1.78%	1.63%	1.68%	1.73%
Other Relatives	9.95%	10.07%	10.19%	9.93%	10.05%	10.18%
Non-Relatives	5.67%	5.76%	5.86%	5.40%	5.50%	5.59%
Group Quarters	2.14%	2.20%	2.25%	2.04%	2.10%	2.16%
SEX:						
Male	50.47%	50.67%	50.86%	50.62%	50.82%	51.03%
Female	49.14%	49.33%	49.53%	48.97%	49.18%	49.38%
RACE:						
White	53.38%	53.60%	53.81%	54.00%	54.23%	54.46%
Black	2.16%	2.23%	2.29%	1.62%	1.68%	1.74%
AIEA	1.03%	1.07%	1.12%	1.23%	1.28%	1.33%
API	3.31%	3.39%	3.46%	4.86%	4.96%	5.06%
Other Races	39.51%	39.72%	39.93%	37.63%	37.85%	38.07%
MARITAL STATUS:						

Now Married	35.01%	35.20%	35.39%	35.72%	35.92%	36.13%
Were Married	10.69%	10.82%	10.94%	10.59%	10.72%	10.85%
Never Married	53.79%	53.98%	54.18%	53.15%	53.36%	53.57%

AGE:

Children (0-18 Years)	37.33%	37.52%	37.71%	36.98%	37.18%	37.38%
Young Adults (19-30 Years)	25.31%	25.48%	25.65%	25.17%	25.36%	25.54%
Adults (31-44 Years)	19.59%	19.75%	19.91%	19.70%	19.87%	20.04%
Mature Adults (45-64 Years)	12.40%	12.53%	12.66%	12.60%	12.74%	12.88%
Elderly Adults (65+ Years)	4.63%	4.72%	4.80%	4.76%	4.85%	4.94%

**Table 11. Comparing Subsets of Hispanic Origin
Against Passel-Word Surname List**

<u>Variable</u>	<u>HISPANIC ORIGIN & SURNAME MATCH Percent</u>	<u>HISPANIC ORIGIN BUT NO MATCH Percent</u>	<u>SURNAME MATCH BUT NO ORIGIN Percent</u>
GEOGRAPHY:			
Mexican Southwest	61.7%	45.6%	42.4%
Puerto Rican Northeast	24.2%	24.6%	19.6%
Florida	5.1%	6.9%	5.2%
Hawaii	0.3%	1.7%	4.4%
Rest of United States	8.7%	21.1%	28.5%
RELATIONSHIP:			
Householder	28.1%	26.2%	26.8%
Husband/Wife	14.2%	19.4%	26.0%
Son/Daughter	38.0%	34.4%	31.2%
Stepchildren	1.7%	2.0%	1.8%
Other Relatives	10.4%	8.8%	6.9%
Non-Relatives	5.6%	6.3%	4.5%
Group Quarters	2.0%	2.9%	2.8%

SEX:

Male	52.1%	45.0%	39.1%
Female	47.9%	55.0%	60.9%

RACE:

White	52.7%	56.8%	65.5%
Black	1.0%	6.8%	7.0%
AIEA	1.0%	1.4%	3.6%
API	2.9%	5.1%	20.1%
Other Races	42.4%	29.9%	3.8%

MARITAL STATUS:

Now Married	35.1%	35.5%	43.1%
Were Married	10.5%	12.1%	12.9%
Never Married	54.4%	52.3%	44.1%

AGE:

Children (0-18 Years)	37.9%	36.2%	31.2%
Young Adults (19-30 Years)	25.3%	26.1%	25.6%
Adults (31-44 Years)	19.5%	20.7%	23.2%
Mature Adults (45-64 Years)	12.6%	12.3%	14.0%
Elderly Adults (65+ Years)	4.7%	4.7%	6.0%

Table 11 emphasizes the PW Spanish surname list's difficulties in accurately detecting Hispanics in traditional Non-Hispanic geographic areas. While 8.7 percent of the Hispanics detected by the PW list come from traditional Non-Hispanic states, 21.1 percent of the Hispanics not detected by the list also come from these states. Additionally, 28.5 percent of the Non-Hispanics whose surnames fall on the PW list comes from these Non-Hispanic states. A review of SOR records with either a Hispanic origin or a matched surname, but not both, might reveal more about the list's low detection rates in these states.

The influence of Native Hawaiians and Filipinos upon the PW Spanish surname list's performance comes out in two categories. First, Hawaii serves as home for 4.4 percent of the Non-Hispanics detected by the PW list, as compared to 0.3 percent of the detected Hispanics. Second, Asians and Pacific Islanders make up 20.1 percent of the Non-Hispanics detected by the PW list, as compared to 2.9 percent of the detected Hispanics. Determining and eliminating the

surnames that occur significantly more often for Filipinos than Hispanics could bring both sets of rates into a smaller range.

Another race that the PW Spanish surname list had difficulties identifying correctly is Blacks. Approximately 1.0 percent of the SOR records with Hispanic origins and surnames found on the PW surname list identify Black as their race. This rate jumps to 6.8 percent for Hispanic SOR records with surnames not found on the list and 7.0 percent for Non-Hispanic SOR records with surnames found on the list. Most Black Hispanics are Puerto Rican, which developed from an African heritage. In their book The Hispanic Population of the United States, Bean and Tienda note that Puerto Rican communities tend to border on Black communities. Such occurrences increases the probability of the intermingling of the two communities and leads to the higher percentages of Blacks containing either a Hispanic origin or a matched PW surname but not both.

The effects of married females taking their husbands' surnames comes across clearly in three of the demographic categories: relationship, sex, and marital status. Spouses encompass 14.2 percent of the Hispanic respondents detected by the PW list, 19.4 percent of the Hispanic respondents not detected by the PW list, and 26.0 percent of the Non-Hispanic respondents that the PW list identified as Hispanic. Females make up 47.9 percent of the Hispanic respondents detected by the PW list, 55.0 percent of the Hispanic respondents not detected by the PW list, and 60.9 percent of the Non-Hispanic respondents the PW list identified as Hispanic. Now married individuals comprise 35.1 percent and 43.1 percent of Hispanics and Non-Hispanics, respectively, identified by the PW list as Hispanic. This increase is offset by the percentages for never married individuals, who comprise 54.4 percent and 44.1 percent of Hispanics and Non-Hispanics, respectively, identified by the PW list as Hispanic.

Although a greater number of records show that Hispanic wives are acquiring Non-Hispanic surnames, the larger percentages occur when Non-Hispanic wives take their husbands' Hispanic surnames. This problem stems directly from the premise of using a surname list to identify Hispanics. Solutions to this problem can come in many forms. Limiting the survey sample to males and never married females eliminates the problem, but such a solution is not practical when gathering data for the general population. Other possible solutions include the use of a Hispanic first name list or the request of a maiden name for married females.

CONCLUSIONS

The PW Spanish surname list appears to detect the Hispanic population as effectively, if not better, as it did in the March 1976 CPS test. Of the respondents in the 1990 Decennial Census SOR file identified by the PW list as Hispanic, 89.98 percent reports a Hispanic origin. The PW list also identifies 79.45 percent of the Hispanic respondents found in the SOR file. Such comparisons of the PW list among different data sets illustrates the list's reliability, as well as possibly any changes in the Hispanic population across time.

In geographic terms, the PW Spanish surname list performs well where it is expected to perform well - in regions with high Hispanic concentration. For the eleven states with high concentrated Mexican, Puerto Rican, and Cuban populations, the PW list provides low surname error rates.

Non-Hispanic respondents comprise 7.59 percent of the respondents whose surnames are on the PW list, while 17.99 percent of the Hispanic respondents did not have matching surnames on the list. For the rest of the United States, the PW Spanish surname list does not perform well. Non-Hispanic respondents comprise 28.91 percent of the respondents whose surnames are on the surname list, while 39.63 percent of the Hispanic respondents did not have matching surnames. Whether these latter surname error rates result because of either the surnames on the PW list or the integration of Hispanics into the American culture remains a matter of further research.

Demographically, the PW Spanish surname list has two major weak points. First, the list has difficulties identifying currently or previously married Hispanic and Non-Hispanic women. The new surname taken by the wife after the marriage may not accurately reflect her Hispanic or Non-Hispanic origin. This problem arises as a general effect of using surname lists to identify a population, not specific to the PW list itself. Second, the Native Hawaiian and Filipino races have genuine Spanish surnames, which results in high SCOM rates for those races and any region that contains a high concentration of them. States that contain a large number of these races include Hawaii and California. Whether or not this is a fixable problem depends on how popular the Spanish surnames shared by Native Hawaiians and Filipinos are among other Hispanics.

Improvement in the PW Spanish surname list can come in a variety of ways. Using the 1990 SOR file, a review of surnames not found on the PW list might reveal a few surnames overwhelmingly reported by Hispanic respondents. Adding the surnames to the PW list would improve its overall SOM rate. A review of surnames on the PW list might find a few surnames rarely reported by Hispanics in the SOR file. Deleting these surnames from the PW list would improve its SCOM rate. A more detailed analysis of areas with numerous cultures may provide information about the popularity of surnames on the PW Spanish surname list among other cultures, most notably, the Native Hawaiian and Filipino cultures. Surnames more popular with these cultures than the Hispanic population should be removed from the PW list. Incorporating an adjoining list of significant Hispanic male and female Hispanic first names, like "Jesus" and "Blanca", with the PW Spanish surname list might identify the correct Hispanic origin for those respondents that change surnames after marriage, decreasing the overall SOM rate. The inclusion of a woman's maiden name before marriage might prove to be a better identifier on this matter. Results generated by undertaking these research projects on the 1990 SOR file will be presented in future Hispanic surname research reports.

REFERENCES

Bean, Frank D. and Marta Tienda. The Hispanic Population of the United States, Russell Sage Foundation: New York, 1987.

Passel, Jeffrey S. and David L. Word. Constructing the List of Spanish Surnames for the 1980 Census: An Application of Bayes' Theorem, paper presented at the Annual Meeting of the Population Association of America, Denver, Colorado, April 10-12, 1980.

U.S. Bureau of the Census. Comparison of Persons of Spanish Surname and Persons of Spanish Origin in the United States, by Edward W. Fernandez, Technical Paper No. 38, June 1975.